AD-A094 056 TEXAS UNIV AT AUSTIN DEPT OF ELECTRICAL ENGINEERING F/G 12/1

ON THE DETERMINATION OF REGRESSION FUNCTIONS,(U)

1979 G L WISE, N C GALLAGHER AFOSR-76-3062

UNCLASSIFIED AFOSR-TR-80-1306 NL

1 OF 1

AD A094056

END

DATE FILMED

2-81

DTIC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-80-1306	2. GOVT ACCESSION NO.: AD-A094056
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): ON THE DETERMINATION OF REGRESSION FUNCTIONS	5. TYPE OF REPORT & PERIOD COVERED: Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Gary L. Wise; Neal C. Gallagher, Jr. (Purdue University)	8. CONTRACT OR GRANT NUMBER(s): AFOSR-76-3062
9. PERFORMING ORGANIZATION NAME AND ADDRESS: The University of Texas at Austin, Department of Electrical Engineering, Austin, Texas 78712	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, D.C. 20332	12. REPORT DATE: October 10-12, 1979
	13. NUMBER OF PAGES: 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE JAN 23 1981 S D E

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Presented at the Seventeenth Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois. Published in the Proceedings of the Conference, pp. 616-623. October 10-12, 1979

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Regression functions
Joint moments

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper is concerned with the determination of regression functions from only a partial characterization of the joint distribution. It is shown that statistical information consisting of various moments and joint moments is sufficient to characterize a regression function. An application to regression functionals is also considered.

DD FORM 1 JAN 73 1473

UNCLASSIFIED 401997

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A094056

AFOSR-TR- 80-1306

ON THE DETERMINATION OF REGRESSION FUNCTIONS

GARY L. WISE
Department of Electrical Engineering
University of Texas at Austin
Austin, Texas 78712

and

NEAL C. GALLAGHER, JR.
School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

This paper is concerned with the determination of regression functions from only a partial characterization of the joint distribution. It is shown that statistical information consisting of various moments and joint moments is sufficient to characterize a regression function. An application to regression functionals is also considered.

I. INTRODUCTION

Let X and Y be random variables with Y integrable, i.e. $E\{|Y|\} < \infty$, and consider the regression function of Y on X,

$$m(x) = E\{Y|X=x\}.$$

As is well known, $m(\cdot)$ is a Borel measurable function, and it frequently arises in engineering applications. For example, if Y is a second order random variable, then the minimum mean squared error estimate of Y in terms of X is given by m(X) [1, pp. 77-78].

In some cases $m(\cdot)$ has a particularly simple form. For example, if X and Y are jointly Gaussian with respective means m_X and m_Y, respective variances $\sigma_X^2 > 0$ and σ_Y^2, and correlation coefficient ρ, then

$$m(x) = ax + b, \tag{1}$$

where $a = (\sigma_Y/\sigma_X)\rho$ and $b = m_Y - am_X$. However, in the case of jointly Gaussian random variables, m_X, m_Y, σ_X, σ_Y, and ρ completely determine the bivariate distribution of the two random variables.

In more general cases, the question arises as to how much information about the bivariate distribution is required to determine the regression function. If X and Y are two second order random variables that are separable in the sense of Nuttall [2], then the regression function $m(\cdot)$ has the form given by (1). However, knowing that two second order random variables are separable in the sense of Nuttall, and knowing the means, variances, and the correlation coefficient is not sufficient to determine the bivariate distribution of the two random variables. Notice that any two random variables whose bivariate characteristic function is elliptically symmetric are separable in the sense of Nuttall [3].

As we have seen, there exists a class of joint distributions such that the regression function can be determined knowing that the two random variables belong to that class and also knowing means, variances, and the correlation coefficient. However, it might seem reasonable to conjecture that in more general cases, the regular conditional distribution [4] of Y

Presented at the Seventeenth Annual Allerton Conference on Communication, Control, and Computing, October 10-12, 1979, to be published in the Proceedings of the Conference.

given X=x is required. Although the regular conditional distribution of Y given X=x is sufficient to determine m(x), in the next section we will see that it is never necessary.

In this paper we will be concerned with statistical information such that there can be only one regression function consistent with the given statistical information. In the next section we consider the regression of Y on a random variable and then on a random vector. Then in the following section we consider the regression functional, that is, the regression of Y on a random process.

II. DEVELOPMENT

Let Y be a second order random variable, let X be a random variable with compact support, and let $m(\cdot)$ be given by Eq. (1). Define the measure μ on the Borel sets of $\mathbb{R}$ by

$$\mu(A) = P(X \in A) \ ,$$

and let $\|\cdot\|$ denote the $L_2(\mu)$ norm. We will say that a polynomial has max degree N if the degree of the polynomial is no greater than N. We note that for any $\varepsilon > 0$, if N is sufficiently large, there exists a polynomial of max degree N $P_N(x)$ such that

$$\| m - P_N \| < \varepsilon. \tag{2}$$

That is, there exists a continuous function $h(\cdot)$ such that [5]

$$\| m - h \| < \varepsilon/2 \ ,$$

and by the Weierstrass Theorem there exists a polynomial P_N of max degree N with N sufficiently large such that

$$\| h - P_N \| < \varepsilon/2 \ .$$

Thus Eq. (2) follows by the triangle inequality. Hence there exists a sequence of polynomials $P_N(x)$ such that

$$P_N(x) \to m(x) \qquad \text{in } L_2(\mu) \ .$$

Let $Q_N(x)$ be the polynomial of max degree N that is closer to m(x) (in $L_2(\mu)$) than any other polynomial of max degree N. We note in passing that $Q_N(x)$ is uniquely defined a.e. $[\mu]$ by the Projection Theorem. That is, there may exist more than one representation of $Q_N(x)$ (i.e. with different coefficients) but they are all equal a.e. $[\mu]$. From the preceding observations, we have that

$$Q_N(x) \to m(x) \qquad \text{in } L_2[\mu] \ .$$

Express the polynomial $Q_N(x)$ as

$$Q_N(x) = \sum_{j=0}^{N} a_j(N)\, x^j \ .$$

It follows from the Projection Theorem that the $a_j(N)$ can be determined from the relation

$$E\left\{\left[m(X) - \sum_{j=0}^{N} a_j(N)\, X^j\right] X^k\right\} = 0, \qquad k = 0, 1, 2, \ldots, N .$$

This is equivalent to

$$E\{X^k Y\} = \sum_{j=0}^{N} a_j(N)\, E\{X^{j+k}\}, \quad k = 0, 1, 2, \ldots, N . \tag{3}$$

Thus we have seen that from a knowledge of

$$E\{X^k\} , \quad k = 1, 2, \ldots$$

and

$$E\{YX^k\} , \quad k = 0, 1, 2, \ldots ,$$

we can construct a sequence of polynomials $Q_N(x)$ that converge in $L_2(\mu)$ to $m(x)$.

Now let X be an arbitrary random variable. Let g be an invertible Borel measurable function whose range is bounded. Define the random variable $\tilde{X}$ as $\tilde{X} = g(X)$, and the measure $\tilde{\mu}$ on the Borel sets of $\mathbb{R}$ by $\tilde{\mu}(A) = P(\tilde{X} \in A)$. From the above discussion, we see that

$$\tilde{m}(x) = E\{Y|\tilde{X} = x\}$$

is determined a.e.$[\tilde{\mu}]$ by the quantities

$$E\{\tilde{X}^k\} , \qquad k = 1, 2, \ldots \tag{4}$$

and

$$E\{Y\tilde{X}^k\} , \qquad k = 0, 1, 2, \ldots . \tag{5}$$

Let $\tilde{Q}_N(x)$ denote the polynomial of max degree N constructed in the preceding fashion. Then

$$\tilde{Q}_N(x) \to \tilde{m}(x) \qquad \text{in } L_2(\tilde{\mu}) .$$

Notice that $m(x) = \tilde{m}[g(x)]$. From a change of variables result [6, p. 182], we have that

$$\int_{g(\mathbb{R})} [\tilde{Q}_N(x) - \tilde{m}(x)]^2\, \tilde{\mu}(dx) = \int_{\mathbb{R}} [\tilde{Q}_N[g(x)] - m(x)]^2\, \mu(dx) .$$

Therefore, $\tilde{Q}_N[g(x)] \to m(x)$ in $L_2(\mu)$.

Now we will remove the restriction that Y be second order. Assume that Y is an integrable random variable and let

$$G_k(y) = \begin{cases} y \text{ if } |y| \le k \\ 0 \text{ if } |y| > k \end{cases} .$$

Then $G_k(Y)$ is a second order random variable and [1, p. 23]

$$E\{G_k(Y)|X=x\} \to E\{Y|X=x\} \quad \text{a.e.}[\mu] .$$

Since $|G_k(Y)-Y| \le |Y|$ and $|Y|$ is integrable, we have that $E\{G_k(Y)|X=x\} \to m(x)$ in $L_1(\mu)$ by the dominated convergence theorem [6, pp. 124-125].

Thus from a knowledge of the quantities in Eqs. (4) and (5) we can derive a sequence of estimates for $E\{G_k(Y)|X=x\}$ which converges in $L_2(\mu)$, and consequently in $L_1(\mu)$ (see, for example, [7]). Also, $E\{G_k(Y)|X=x\}$ converges to $E\{Y|X=x\}$ in $L_1(\mu)$. Thus, by a straightforward diagonalization procedure, we can derive a sequence of estimates which converges in $L_1(\mu)$ to $m(x)$. These results are summarized in the following theorem.

Theorem 1: Let Y be an integrable random variable, let X be an arbitrary random variable, and let g be an invertible Borel measurable function mapping the reals into a bounded set. Then the regression function m is determined a.e.[μ] by the quantities

$$E\{[g(X)]^k\}\ , \quad k = 1, 2, \ldots$$

and

$$E\{Y[g(X)]^k\}\ , \quad k = 0, 1, 2, \ldots\ .$$

Consider for the moment the case where X and Y are independent. In this case a solution to Eq. (3) is given by

$$a_0(N) = E\{Y\}$$

$$a_j(N) = 0\ , \ j > 0\ ,$$

and we get that $m(x) = E\{Y\}$.

Now consider the following two different bivariate density functions:

$$f_1(x,y) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(y-\rho x)^2}{2\sigma^2}\right] I_{[0,1]}(x)$$

$$f_2(x,y) = x I_{[\rho x-1,\rho x+1]}(y)\ I_{[0,1]}(x),$$

where $\sigma > 0$, $\rho \in (-1,1)$, and I denotes the indicator function. Assuming that the density of (X,Y) is given by f_1, we find that

$$E\{X^k\} = \frac{1}{k+1}$$

$$E\{YX^k\} = \frac{\rho}{k+2}\ .$$

In this case, for $N \geq 1$, a solution to Eq. (3) is given by

$$a_1(N) = \rho \tag{6}$$

$$a_j(N) = 0\ , \ j \neq 1\ , \tag{7}$$

and we conclude that

$$m(x) = \rho x\ . \tag{8}$$

If we assume that the density of (X,Y) is given by f_2, we find that

$$E\{X^k\} = \frac{2}{k+2}$$

$$E\{YX^k\} = \frac{2\rho}{k+3} .$$

In this case, for $N \geq 1$, Eqs. (6) and (7) still satisfy Eq. (3), and the regression function is once again given by Eq. (8). Thus, in this example, the two pairs of marginal densities are not the same, the conditional densities of Y given X=x are not the same, and the moment sequences are not the same; however, the moment sequences are sufficient to characterize the conditional expectations, which are identical. Numerous other similar examples may easily be constructed.

Now we will consider the regression of Y upon a set of random variables. Let X be an arbitrary random vector taking values in $\mathbb{R}^n$, and let μ be defined on the Borel sets of $\mathbb{R}^n$ by

$$\mu(B) = P(X \in B) .$$

Lemma 1: If μ has compact support, then the class of all polynomials is dense in $L_2(\mu)$.

Proof: Let q be an arbitrary element in $L_2(\mu)$. For any $\varepsilon > 0$, there exists [5] a function $h: \mathbb{R}^n \to \mathbb{R}$ which is continuous and has compact support such that

$$\| q-h \| < \varepsilon/2 .$$

By the Stone-Weierstrass Theorem [8] there exists a polynomial p in n variables such that

$$\| h-p \| < \varepsilon/2 ,$$

and thus by the triangle inequality

$$\| p-q \| < \varepsilon .$$

QED

We recall that the degree of a monomial in n variables is the sum of the powers of the variables, and the degree of a polynomial is the degree of the monomial having the largest degree over all the monomials in the polynomial with nonzero coefficients. There are

$$C(n,d) = \binom{n+d-1}{d}$$

monomials of degree d in n variables [9].

Assume that Y is a second order random variable, and define m(x) by Eq. (1), where x is now an element of $\mathbb{R}^n$. Assume that μ has compact support. Let $Q_N(x)$ be the polynomial of max degree N which is closer, in the $L_2(\mu)$ norm, to m(x) than any other polynomial of max degree N.

Consider a monomial in n variables of degree d. There will be C(n,d) of them. Order them lexicographically by the powers of the components of x, and let $m_{jd}(x)$ denote the j-th monomial of degree d. Then $Q_N(x)$ can be expressed as

$$Q_N(x) = \sum_{d=0}^{N} \sum_{j=1}^{C(n,d)} a_{jd}(N) m_{jd}(x) \ .$$

It follows from the Projection Theorem that the coefficients $a_{jd}(N)$ are given by the solution to the following set of equations:

$$E\{Ym_{ik}(X)\} = \sum_{d=0}^{N} \sum_{j=1}^{C(n,d)} a_{jd}(N)\ E\{m_{jd}(X)\ m_{ik}(X)\}, \tag{9}$$

$k = 0, 1, \ldots, N$ and $i = 1, \ldots, C(n,k)$. If the coefficients $a_{jd}(N)$ satisfy Eq. (9), then it follows from Lemma 1 that

$$Q_N(x) \to m(x) \qquad \text{in } L_2(\mu) \ .$$

Now we remove the assumption that X has compact support and let X be an arbitrary random vector taking values in $\mathbb{R}^n$. Let g be an invertible Borel measurable function mapping $\mathbb{R}^n$ into a bounded subset of $\mathbb{R}^n$, and let $\tilde{X} = g(X)$. We see that

$$\tilde{m}(x) = E\{Y|\tilde{X}=x\}$$

is determined a.e.$[\tilde{\mu}]$, where $\tilde{\mu}(A) = \mu[g^{-1}(A)]$, by the quantities

$$E\{m_{jd}(\tilde{X})\}$$

and

$$E\{Ym_{jd}(\tilde{X})\}$$

for $d = 0, 1, 2, \ldots$ and $j = 1, \ldots, C(n,d)$. Let $\tilde{Q}_N(x)$ be the polynomial of max degree N determined in the preceding fashion. Then, similar to the development of Theorem 1, we can employ a change of variables result [6, p. 182] to conclude that

$$\tilde{Q}_N[g(x)] \to m(x) \qquad \text{in } L_2(\mu) \ .$$

A chopping argument as in the development of Theorem 1 allows us to remove the second order restriction on Y. Then a straightforward diagonalization procedure results in a sequence of estimates which converges to m(x) in $L_1(\mu)$. This result is summarized in the following theorem.

<u>Theorem 2</u>: Let Y be an integrable random variable, let X be an arbitrary random vector taking values in $\mathbb{R}^n$, and let g be an invertible Borel measurable function mapping $\mathbb{R}^n$ into a bounded subset of $\mathbb{R}^n$. Then the regression function m is determined a.e.$[\mu]$ by the quantities

$$E\{m_{jd}[g(X)]\} \quad \text{and} \quad E\{Ym_{jd}[g(X)]\}$$

for $d = 0, 1, 2, \ldots$ and $j = 1, \ldots, C(n,d)$.

III. REGRESSION FUNCTIONALS

As before, assume that Y is an integrable random variable, but now let T be an infinite subset of $\mathbb{R}$ and let $\{X(t),\ t \in T\}$ be a random process. Let S denote the space of all extended real valued functions defined on T, and let $\mathscr{B}(S)$ denote the σ-algebra on S generated by the class of all cylinders in S. Let $\mathscr{B}$ denote the Borel sets of $\mathbb{R}$. Then the regression functional

$$m[x(t),\ t \in T] = E\{Y|X(t) = x(t),\ t \in T\}$$

is a measurable function from $(S, \mathscr{B}(S))$ to $(\mathbb{R}, \mathscr{B})$ (see, for example, [10]).

Let μ be the measure induced on $\mathscr{B}(S)$ by $\{X(t),\ t \in T\}$. That is, for any cylinder C in S, $\mu(C) = P(\{X(t),\ t \in T\} \in C)$, and μ is extended to $\mathscr{B}(S)$ via Kolmogorov's Theorem (see, for example, [11]).

It follows from [1, pp. 21, 604] that there exists a countable subset of T, say $\tilde{T} = \{t_1, t_2, \ldots\}$, depending on the random variable Y, such that

$$E\{Y|X(t) = x(t),\ t \in T\} = E\{Y|X(t) = x(t),\ t \in \tilde{T}\} \text{ a.e.}[\mu] \ .$$

Let

$$M = E\{Y|X(t),\ t \in \tilde{T}\},$$

$$M_n = E\{Y|X(t_1),\ \ldots,\ X(t_n)\},$$

$$\mathscr{F} = \sigma\{X(t),\ t \in \tilde{T}\},$$

and

$$\mathscr{F}_n = \sigma\{X(t_1),\ \ldots,\ X(t_n)\}.$$

Then from the properties of iterated conditional expectations [1, p. 37], it follows that

$$E\{M_{n+1}|\mathscr{F}_n\} = M_n \quad \text{wp1} \ ,$$

and hence $\{M_n, \mathscr{F}_n, n \geq 1\}$ is a martingale. It follows from [1, p. 332] that $M_n \to M$ wp1. Since $E\{|M_n|\} \leq E\{|Y|\} < \infty$, it follows from a martingale convergence theorem [1, p. 319] due to Doob that $E\{|M_n - M|\} \to 0$. This is equivalent to

$$E\{Y|X(t_i) = x(t_i),\ i=1,\ldots,n\} \to E\{Y|X(t) = x(t),\ t \in \tilde{T}\}$$

in $L_1(\mu)$. Notice that Theorem 2 is applicable to $E\{Y|X(t_i) = x(t_i),$ $i=1,\ldots,n\}$. Thus a straightforward diagonalization procedure results in a sequence of estimates which converges to $m[x(t),\ t \in T]$ in $L_1(\mu)$. This result is summarized in the following theorem.

<u>Theorem 3</u>: Let Y be an integrable random variable and let $\{X(t),\ t \in T\}$ be a random process. Let $\{g_n,\ n=1,2,\ldots\}$ be a sequence of functions where g_n is an invertible Borel measurable function from $\mathbb{R}^n$ to a bounded subset of $\mathbb{R}^n$. Assume that for all positive integers n and for all sets

of n points in T, say $t_1, \ldots, t_n$, the quantities

$$E\{m_{jd}(g_n[X(t_1), \ldots, X(t_n)])\}$$

and

$$E\{Ym_{jd}(g_n[X(t_1), \ldots, X(t_n)])\}$$

for $d = 0, 1, 2, \ldots$ and $j = 1, \ldots, C(n,d)$ are known. Then up to μ equivalence, there is only one possible regression functional $m[x(t), t \in T] = E\{Y|X(t) = x(t), t \in T\}$.

ACKNOWLEDGEMENT

This research was supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grants AFOSR-76-3062 and AFOSR-78-3605.

REFERENCES

1. J. L. Doob, Stochastic Processes, Wiley, New York, 1953.

2. A. H. Nuttall, "Theory and Application of the Separable Class of Random Processes," Technical Report 343, Research Laboratory of Electronics, Massachusetts Institute of Technology, May 26, 1958.

3. D. K. McGraw and J. F. Wagner, "Elliptically Symmetric Distributions," IEEE Trans. Inform. Th., Vol. IT-14, pp. 110-120, January 1968.

4. L. Breiman, Probability, p. 79, Addison-Wesley, Reading, Mass., 1968.

5. W. Rudin, Real and Complex Analysis, p. 71, McGraw-Hill, New York, 1974.

6. N. Dunford and J. T. Schwartz, Linear Operators Part I: General Theory, Interscience, New York, 1957.

7. M. Loève, Probability Theory, p. 164, Van Nostrand, New York, 1963.

8. J. Dieudonné, Foundations of Modern Analysis, p. 139, Academic Press, New York, 1969.

9. R. W. Brockett, "Lie Algebras and Lie Groups in Control Theory," in Geometric Methods in System Theory, D. Q. Mayne and R. W. Brockett, eds., Reidel, The Netherlands, 1973, pp. 43-82.

10. I. I. Gihman and A. V. Skorohod, The Theory of Stochastic Processes I, p. 34, Springer-Verlag, New York, 1974.

11. P. Billingsley, Probability and Measure, p. 433, Wiley, New York, 1979.

DATE
FILMED